Naval Research Laboratory

Washington, DC 20375-5000

NRL Report 9305

A Logic for the Analysis of Cryptographic Protocols

PAUL SYVERSON

Center for Secure Information Technology
Information Technology Division

December 31, 1990

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 31, 1990	Oct. 1989-Sept. 1990

4. TITLE AND SUBTITLE

A Logic for the Analysis of Cryptographic Protocols

5. FUNDING NUMBERS

TA - RR015-03-41
WU - 2829-0-1-(6.1)

6. AUTHOR(S)

Syverson, Paul F.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL Report 9305

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Arlington, VA 22217-5000

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Some of the results presented in this report were previously published in the Computer Security Foundations Workshop III, Franconia, New Hampshire.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A logic designed to analyze cryptographic protocols is presented in this report. The logic has distinct means for representing propositional knowledge and knowledge in the sense of familiarity with an individual, e.g., a particular key. It is argued that the introduction of a knowledge predicate is useful and genuine increase in expressive power. The semantics and metalogic of the logic are also explored.

14. SUBJECT TERMS

Computer security
Formal methods
Protocol analysis
Logic

15. NUMBER OF PAGES

20

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

CONTENTS

A LOGIC FOR THE ANALYSIS OF CRYPTOGRAPHIC PROTOCOLS

INTRODUCTION

In this report, we present a free epistemic logic with separate means for explicitly representing both propositional knowledge and knowledge of individuals. The logic has been designed primarily for the analysis of cryptographic protocols, but it is not necessarily limited to this application. Thus, the logic has distinct mechanisms for representing—e.g., knowing that k is Saul's crypto-key vs knowing k in the sense of being able to recognize or produce it. These representations are accomplished by means of a standard knowledge operator and a knowledge predicate respectively. The logic presented here is the result of a significant revision of the logic given in (Syverson, 1990). The current version corrects certain errors and omissions in the original account. We also argue briefly that the introduction of a knowledge predicate is more than mere novelty; it facilitates a genuine and valuable expansion of expressive power.

In "The Use of Logic in the Analysis of Cryptographic Protocols" (submitted for publication), we argue that it is valuable for a crypto-protocol logic to have an independently motivated semantics, one that explicitly incorporates the cryptographic features of the logic. In addition to other advantages, if the logic is shown to be sound and complete with respect to the semantics, then we have strong assurance that the logic captures all and only the valid reasoning expressible in the formal language. One of the primary goals of this report is to present such metalogical results. Before proving these, however, we set out the language, semantics, and logic.

THE LANGUAGE

The language contains a denumerable number of names of words: s_1, s_2, s_3, ... Each word should be thought of as a string of symbols from some finite alphabet, e.g., a key. However, since we need not depict the structure of words in our language, they are represented atomically. The language also contains equality and two functions taking pairs of words to words.[1] $e(x, y) = z$ should be taken to mean that z is the result of encrypting y with key x. $d(x, y) = z$ should be taken to mean that z is the result of decrypting y using key x. (x, y, z, ... are variables ranging over arbitrary words.) Our language also contains denumerably many predicate constants, each of finite arity and taking tuples of word names as arguments: P_1, P_2, P_3, ... Of these we call particular attention to a set of unary epistemic predicate constants: C_1, ..., C_n. Intuitively $C_i(x)$ should be taken to mean that i knows x, i.e., i can recognize or produce the character string named by x. If i is able to decrypt a message he receives that has been encrypted with key x, this serves as evidence

Manuscript approved October 2, 1990.

1. To be precise we should say that the language contains the identity symbol and two function symbols that represent equality and two functions respectively, but we adopt common use-mention confusions where it is harmless.

that $C_i x$ is true; he must have known the key since he was able to use it. (It is assumed that we are talking about symmetric encryption here, i.e., the encryption and decryption keys are the same.)

The language is first order with quantification being over words. The basic (open) sentences of the language are expressions of the form $P_k(x_1, ..., x_j)$ or $P_k(s_1, ..., s_j)$ and equations. Closed sentences are those containing no free variables. Sentences (open or closed) may be assembled into (finite) complex sentences according to ordinary recursive formation rules using the usual connectives: $\neg$, $\wedge$, $\vee$, and $\rightarrow$. The only remaining feature of the language is a finite set of propositional knowledge operators: $S_1, ..., S_n$. These are standard epistemic operators in the style of Hintikka (1962). Intuitively $S_i\varphi$ should be taken to mean that i knows the proposition expressed by φ.[2] (φ is a variable ranging over arbitrary sentences.) $S_i\varphi$ is a sentence, provided that φ is a sentence. Thus these operators may be iterated, although we will not have need to do so in this report. Note the difference between C_i and S_i. C_i is a predicate; it applies to words (individuals). S_i is an operator; it applies to sentences.

SEMANTICS

The semantics we adopt is a slight modification of the standard Hintikka style possible world semantics for epistemic logics. Before setting things out formally we will give an intuitive picture. First we have a set of possible worlds. These may be thought of as all the different ways the world may be. On this set there is an accessibility relation between worlds for each individual i. If world w′ is accessible from w for a given individual, then that individual in w cannot distinguish the two worlds given his current state of knowledge. Thus, suppose there are two worlds that are accessible to each other for i. In one of these worlds it is raining, and in the other it is not. In this case, i does not know whether or not it is raining (relative to either world). If a sentence φ is true in all the worlds accessible for i from some world w_n, then we can say that i knows φ in that world. N.B. φ may actually be false! This is because we have said nothing about how w_n compares to the actual world. While φ may be false in the actual world, if it is true in w_n and in all worlds accessible from w_n for i, then i knows φ in w_n. Now, we are usually worried about what someone knows in the actual world. So, "i knows φ." (*simpliciter*) should be taken to mean that i knows φ in the actual world.

The above corresponds to our characterization of propositional knowledge by means of the S_i operators. For the knowledge characterized by the C_i predicates, we maintain the same semantic structure of worlds and accessibility relations; we simply add to it. In quantified modal logic one decides whether, for example, $P(x_1, ..., x_k)$ is true at a world by seeing if the k-tuple of values assigned to x_1 through x_k respectively at that world is in the set assigned to P at that world. The same criterion applies to sentences formed with the C_i predicates. This is somewhat unusual; except for identity, predicates usually receive their interpretation extralogically. The interpretation of C_i is intimately tied to the semantic structure itself. $C_i x$ is true at a world w_n whenever x is assigned a value at w_n and it is assigned the same value at all worlds accessible from w_n for i.

2. The choice of symbols for knowledge derives from the French words '*connaître*' and '*savoir*'. For example, in French, you *connais* a person and you *sais* that it's raining. In English, which does not make the distinction, both of these mean to know.

Since these predicates are unusual, we shall give a little explanation of their semantic interpretation.

The possible worlds represent the different ways someone thinks reality might be. If world w′ is accessible from world w for some subject, e.g. Scott, then at world w, he cannot tell them apart. From the perspective of world w, Scott finds both w and w′ equally possible ways things might be. Now, suppose some word s is present at one of these worlds but not the other. (What 'present' means will be clearer once the model theory is spelled out below.) Then Scott cannot tell the difference between a world where s is present and one where it is not. So, he must not really be aware of the word, *know* the word, if he can't tell whether it's there or not. Under these circumstances we would not want to say that he can recognize or reproduce the word. Thus it should indeed turn out that $C_{Scott}(s)$ is not true at w.

Domains, Terms, and Denotations

There is a potential problem with our semantics. If every term of the language were to denote in every world, and if terms always denoted the same word regardless of the world, then everyone would know all the words in all circumstances—assuming all the words were named in the language. This is so because all the worlds would have the same words in them, and those words would be named the same way at each of them. This would render the C predicates trivial and thus useless. The answer of course is to vary the domain of quantification from world to world. This will block the validity of $\forall x\, C_i x$ as long as there are things in the domain of quantification of some world that are not in the domain of quantification in another.[3] Unfortunately, this strategy is not sufficient to entirely solve the problem. For, even with the domains varying, a constant term will (by definition) denote the same word in all worlds. Thus, any word that is given a name in our language will be a word that everyone always knows. Somehow we need to have terms that may not denote in all possible worlds. Fortunately, there is a way to deal directly with nondenoting singular terms.

Free Logic

Ermanno Bencivenga (1986) defines a free logic as "a formal system of quantification theory, with or without identity, which allows for some singular terms in some circumstances to be thought of as denoting no existing object, and in which quantifiers are invariably thought of as having existential import."[4] This is just what we want, provided that we fill in the details properly.

In effect, the strategy here is to adopt the proposal given above, namely to vary the domain of quantification from world to world. All we need do is incorporate the correct interpretation of terms into this picture. A singular term t denotes at a world just in case it names a member of the domain of quantification at that world, i.e., $\exists x\, (x = t)$ is true at that world (x is a variable distinct from t). For ease of expression, we define a predicate expressed by 'E' such that

3. Note that this also provides a semantic guarantee that the Barcan Formula is not valid. We will return to this below where it will be seen to be a desirable result.

4. *op. cit.*, p. 375.

$E(t) =_{df} \exists x\ (x = t)$ —where x is a variable distinct from t. In place of the classical quantifier rules, we have the following.

Universal Instantiation

From $\forall x\ \varphi \wedge Et$	for any term t, where φ is a sentence in the language, and
infer $\varphi[t/x]$	$\varphi[t/x]$ is the same sentence as φ except that all free occurrences of x in φ are replaced by t

Universal Generalization

From $\psi \rightarrow (Et \rightarrow \varphi)$	where t is a term that does not occur freely in ψ or in
infer $\psi \rightarrow \forall x\ \varphi[x/t]$	any assumption on which $\psi \rightarrow (Et \rightarrow \varphi)$ depends

Intuitively understood, these may help in comprehending what it means for quantification to always have existential import. We need to spell out formal models and interpretations to see exactly how these rules work, and that is what we do now.

Models

A model is a tuple $<W, R_1, ..., R_k, D, d, a>$ where W is a set of nonempty possible worlds, $R_1, ..., R_k$ are binary accessibility relations between members of W, and D is a domain of objects for all possible worlds. d is a function from members of W to subsets of D, thus d(w) is the domain at world w. a is an assignment function, which assigns values to expressions in the language in the manner given below. Since we want to allow a to be undefined sometimes, we adopt the standard trick of adding a value $*$ to represent being undefined. This allows us to have an assignment function that is total and yet still gives us a means to say that terms sometimes fail to denote and sentences sometimes do not have a definite truth value. Note that since an assignment function does the duty of both an interpretation and a valuation, $*$ can do the duty of both an undefined truth value and an undefined member of a domain.

$a(t) \in D$	for all terms t
$a(<t_1, ..., t_n>) = <a(t_1, ..., a(t_n)>$	where $t_1, ..., t_n$ are terms (names of words)

(We suppress tuple notation from here on when it is clear what is meant.)

$a(f(t_1, ..., t_n)) = a(f)(a(t_1, ..., t_n)) =$ $= a(f)\ (a(t_1, ..., a(t_n))$	where $t_1, ..., t_n$ are terms (names of words) and f is the name of a function on words
$a(P)$ is a set of n-tuples of members of D	where P is any n-ary predicate letter ($n \geq 1$)

Encryption and decryption pose a problem for an assignment function; neither the encryption nor the decryption key is necessarily unique. For example, in the RSA algorithm any power of a key is also a key, i.e., something encrypted using a power of the encryption key can be decrypted using the usual decryption key, and vice versa. This may also be true of symmetric en-

cryption schemes. Informally we shall follow the conventional pretense that encryption and decryption keys are unique whenever such pretense causes no harm.[5] Thus, if k is the name of some key, k^{-1} is to be intuitively interpreted as the name of the inverse key corresponding to k. In order to ensure that all works out properly on the formal level we define the following.

For a given key (term) k,

$[k] = \{t : a(e(t, y)) = a(e(k, y)) \text{ for each } y \in a^{-1}(D)\}$

$[k^{-1}] = \{t : a(d(t, e(k, y))) = a(y) = a(e(k, d(t, y))) \text{ for each } y \in a^{-1}(D)\}$

a_w is the restriction of a to d(w) on the above type arguments and also satisfying the following.

$a_w(t)$ is

- $a(t)$ if $a(t) \in d(w)$
  - or if $a(t) = a(f(t_1, ..., t_n))$ for some $t_1, ..., t_n$ s.t. $a(t_1), ..., a(t_n) \in d(w)$
  - or if $a_w(e(s_1, s_2)) = a(e(k, t))$ for some s_1, s_2 s.t. $a(s_1), a(s_2) \in d(w)$ and some k s.t. $k' \in [k^{-1}]$ and $a(k') \in d(w)$
  - or if $a_w(d(s_1, s_2)) = a(d(k', t))$ for some s_1, s_2 s.t. $a(s_1), a(s_2) \in d(w)$ and some k' s.t. $k' \in [k^{-1}]$, $a(k) \in d(w)$
- $*$ otherwise

This is not as complicated as it looks. There are four cases under which a term t denotes at a world w. The first case is when it is simply given. Perhaps t is a public key that everyone knows, thus it is present at every world. The second case is when t names the same thing as a function of terms, and each of the arguments of the function denotes at w. The third case is when e(k, t) is assigned the same value as a word that is an encrypted word at w, and the decryption key also denotes at w. It is important to note that it is not enough that a(e(k, t)) = a(s) for some s that denotes at w. s must be an encrypted word in w, not just in D. Intuitively, in order to apply a decryption key to a word in world w, that word must be an encrypted word *in w*. The fourth case is similar to the third except that it deals with decrypted words rather than encrypted words.

$a_w(s = t)$ is

- T if $a_w(s) = a_w(t)$ and $a(s) \in d(w)$
- F if $a(s) \neq a(t)$ and $a(s), a(t) \in d(w)$
- $*$ otherwise

$a_w(C_i t)$ is

- T if $a(t) \in d(w')$ for all w' such that wR_iw'
- F if $a(t) \in d(w)$ and $a(t) \notin d(w')$ for some w' s.t. wR_iw'
- $*$ otherwise

5. For convenience, we also restrict ourselves to cryptosystems with two sided inverses. This is not a serious restriction as it covers those cryptosystems that are currently in widest use.

For n-ary predicate letters P, other than equality and C_i (for i = 1, ..., k), we have

$$a_w(P(t_1, ..., t_n)) \text{ is } \begin{cases} T & \text{if } a(t_1), ..., a(t_n) \in d(w) \text{ and } a(t_1, ..., t_n) \in a(P) \\ F & \text{if } a(t_1), ..., a(t_n) \in d(w) \text{ and } a(t_1, ..., t_n) \notin a(P) \\ * & \text{otherwise} \end{cases}$$

For an arbitrary sentence φ,

$$a_w(S_i\varphi) \text{ is } \begin{cases} T & \text{if } a_{w'}(\varphi) = T \text{ for all } w' \text{ such that } wR_iw' \\ F & \text{if } a_{w'}(\varphi) \text{ is defined for all } w' \text{ such that } wR_iw' \\ & \text{and } a_{w'}(\varphi) = F \text{ for some } w' \text{ such that } wR_iw' \\ * & \text{otherwise} \end{cases}$$

$$a_w(\forall x\, \varphi) \text{ is } \begin{cases} T & \text{if } a_w(\varphi[t/x]) = T \text{ for all t such that } a(t) \in d(w) \\ F & \text{if } a_w(\varphi[t/x]) = F \text{ for some t such that } a(t) \in d(w) \\ * & \text{otherwise} \end{cases}$$

where $\varphi[t/x]$ is the same sentence as φ except that all free occurrences of x in φ are replaced by t

$$a_w(\varphi \wedge \psi) \text{ is } \begin{cases} T & \text{if } a_w(\varphi) = T \text{ and } a_w(\psi) = T \\ F & \text{if } a_w(\varphi) = F \text{ or } a_w(\psi) = F \\ & \text{and both } a_w(\varphi) \text{ and } a_w(\psi) \text{ are defined} \\ * & \text{otherwise} \end{cases}$$

$$a_w(\varphi \vee \psi) \text{ is } \begin{cases} T & \text{if } a_w(\varphi) = T \text{ or } a_w(\psi) = T \\ & \text{and both } a_w(\varphi) \text{ and } a_w(\psi) \text{ are defined} \\ F & \text{if } a_w(\varphi) = F \text{ and } a_w(\psi) = F \\ * & \text{otherwise} \end{cases}$$

$$a_w(\varphi \rightarrow \psi) \text{ is } \begin{cases} T & \text{if } a_w(\varphi) = F \text{ or } a_w(\psi) = T \\ & \text{and both } a_w(\varphi) \text{ and } a_w(\psi) \text{ are defined} \\ F & \text{if } a_w(\varphi) = T \text{ and } a_w(\psi) = F \\ * & \text{otherwise} \end{cases}$$

$$a_w(\neg\varphi) \text{ is } \begin{cases} T & \text{if } a_w(\varphi) = F \\ F & \text{if } a_w(\varphi) = T \\ * & \text{if } a_w(\varphi) = * \end{cases}$$

KNOWLEDGE REPRESENTATION

Now that the basic linguistic and semantic structures are in place, we can say something about the mechanisms for knowledge representation. One important question is whether or not we need the knowledge predicates (C predicates) as primitives at all. Is there not some way that we can define them in terms of the knowledge operators? Obvious candidates for defining $C_i t$ are

$S_i(Et)$ and $\exists x\, S_i(x = t)$. Indeed these are the standard ways of representing this type of knowledge. Debate on which of these is more appropriate is forestalled by their equivalence in the above semantics and is a fortiori precluded once we realize that neither can ever be false. In any given world they are both either true or undefined. Nonetheless, perhaps this indicates not the indispensability of the knowledge predicates, but the inadequacy of the assignment function.

Perhaps the assignment function makes unnecessary distinctions in the case of the knowledge predicates. The way the assignment function works for the C predicates does seem a little odd. Recalling the possible worlds explanation of them given above may clarify the reasons for the assignment of T, but what about the distinction between being assigned F and being undefined? Obviously an atomic sentence that contains a term that fails to denote at a world should fail to have a truth value at that world. (Recall that ours is a logic of epistemic, not alethic[6], modalities. If 'Pegasus' does not denote at a world, it doesn't mean simply that he does not exist there; it means that he is not known there.) This explains the conditions under which $a_w(C_it)$ is undefined. As for falsity, from i's perspective C_it should never be false. How could i think that he does not know a word (in the C sense)? To think about the word at all he must know what it is. But, others may be in a position to realize that i does not know a word, and there may be things that follow logically from the falsity of C_it even if no subject knows it. So, we must be able to assign the value F to C_it. We have already seen that t must denote for C_it to be assigned a truth value at all, and if t were to denote in all worlds accessible from w for i, then C_it would clearly be true. Thus, the only way for C_it to be assigned the value F at a world w is if t denotes at w but fails to denote at some world accessible from w for i.

This justification still does not ensure the necessity of primitive knowledge predicates. Perhaps it is the semantics of the knowledge operators that must be changed, and once this is done correctly, the predicates will be reducible to the operators. We could redefine the assignment of truth values to the knowledge operators so that, for example, $S_{Scott}\varphi$ is false at w if and only if it is defined at w and false or undefined at some world accessible from w for Scott. This would make C_it and $S_i(Et)$ semantically equivalent. Unfortunately such a move would obliterate the distinction between Scott's knowing that φ is true and his recognizing it as meaningful. This distinction is important to the evaluation of cryptographic protocols, the primary purpose for which this logic was devised. For example, let us suppose that the security of a protocol we are evaluating depends on the secrecy of Louie's key k. The protocol should be secure enough if we can conclude that penetrator Scott does not know that k is Louie's key. But, it is still more secure if Scott does not even know that "k is Louie's key." is meaningful. At the very least there is a difference between these two situations of how much additional information the penetrator must obtain to render the protocol insecure. Thus, from a semantic point of view, the knowledge predicates are both useful and noneliminable.

THE LOGIC

It should be clear from the language set out above that the logic we are about to present will be a quantified modal logic. These are notoriously difficult semantically. In addition to the problems associated with modality per se there are a number of problems associated with the in-

6. Alethic modalities are the modalities of necessity and possibility.

teraction of modality and quantification.[7] We intend to skirt as many of the issues as we can that do not bear directly on the subject of this report. For instance, we have chosen the logic **T** as the basic epistemic logic because we have no need to represent iterated propositional knowledge. However, this should not be viewed as a commitment to a position on introspection. In applications where such needs might arise I would be perfectly willing to use other logics, such as **S4** or **S5**, if this did not create any problems. (We will see below that we must reject both the Barcan Formula and its converse, thus **S5** is already ruled out.)

'Standard' Axioms and Rules

Axioms 1 through 5 are the universal closures of the following, where there are no freely occurring constant terms in α, β, or γ.

1. $\alpha \rightarrow (\beta \rightarrow \alpha)$
2. $(\alpha \rightarrow .\beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \beta \rightarrow .\alpha \rightarrow \gamma)$
3. $(\neg\beta \rightarrow \neg\alpha) \rightarrow (\neg\beta \rightarrow \alpha \rightarrow \beta)$
4. $S_i\,\alpha \wedge S_i\,(\alpha \rightarrow \beta) \rightarrow S_i\,\beta)$
5. $S_i\,\alpha \rightarrow \alpha$

The reason for the restrictions on axioms 1 through 5 is to make sure that they are true in all models. Without the restrictions, an axiom would not have a defined truth value at a world if it contained a term that failed to denote there. While the idea of axioms that are not necessarily true at all worlds is somewhat bizarre, there is no harm in it; however, for convenience and to avoid unnecessary confusion, we adopt the above restrictions.

6. $\forall x\,(x = x)$
7. $\forall x\,\forall y\,(x = y \rightarrow .\varphi \rightarrow \varphi')$

   (where φ' is the result of placing no, some, or all occurrences of 'x' in φ with 'y', and where neither φ nor φ' contain any free occurrences of any constant terms)

There are also two rules of inference:

R1. From φ and $\varphi \rightarrow \psi$ infer ψ (Modus Ponens)

R2. From $\vdash \varphi$ infer $\vdash S_i\varphi \quad i = 1, ..., n$ (Epistemic Generalization)

φ and ψ may be either open or closed in modus ponens but not in epistemic generalization. (The reason for the restriction to closed sentences in this case is explained below.) Even these basic axioms and rules are problematic. The axioms together with R2 yield the omniscience problem; each subject knows all logical truths. Various attempts have been made to solve this and other related problems by restricting the logic in one way or another (Eberle 1974; Fagin and

7. For an analysis of some of the major issues c.f. (Garson 1984).

Halpern 1985; Levesque 1983). Other research has been done on non-monotonic doxastic logic for computer security in which posited beliefs may be taken back (Moser 1989). Still others have analyzed complexity issues in reasoning about knowledge and belief (Goldwasser et al. 1985; Halpern and Vardi 1986). The difficulties these papers deal with are serious problems and not just for the correct theoretical representation of reasoning. From a practical standpoint, in computer security we don't want to waste time worrying about inferences that no penetrator will ever actually draw. We also don't want to be overly confident about what we ourselves can discern about a penetrator.

Despite these problems, logics that model reasoning without restrictions of complexity have been quite useful in uncovering important properties of distributed systems and of cryptographic protocols. And, there is invariably a trade-off between the accuracy gained by less idealized analyses and the ease and speed with which such analyses are done. This trade-off is all the more pronounced if the idealized system has associated semantic techniques available. So, while we acknowledge this problem, we do not attempt to deal with it here.

Because of the omniscience problem, it is perhaps wrong or at least misleading to interpret the S_i's and C_i's epistemically. We have done so partly to maintain terminological consistency with previous work and partly because that work is not so far off. Jon Barwise has said that "information travels at the speed of logic, genuine knowledge travels only at the speed of cognition and inference," and that "much of the work in the logic of knowledge is best understood in terms of the logic of information." (Barwise 1989, p. 204) I am entirely in agreement with these sentiments. Consequently, $S_i\varphi$ is probably more accurately understood as saying that i has information that φ. Similarly, $C_i x$ is probably best understood as saying that i has sufficient information to recognize x or to produce it. Despite these points, we will retain the terminology we started with for the remainder of the report.

Rules for Quantification and for Relating Types of Knowledge

The basic quantifier rules are what distinguish this as a free logic—as opposed to a classical one. The rules were introduced above, and we restate them here as official rules of the logic. $\exists x$ is defined as $\neg\forall x\neg$ as usual.

R3. (Universal Instantiation)

From $\forall x\ \varphi \wedge Et$	for any term t, where φ is a sentence in the language, and
infer $\varphi[t/x]$	$\varphi[t/x]$ is the same sentence as φ except that all free occurrences of x in φ are replaced by t

R4. (Universal Generalization)

From $\psi \rightarrow (Et \rightarrow \varphi)$	where t is a term that does not occur freely in ψ or in
infer $\psi \rightarrow \forall x\ \varphi[x/t]$	any assumption on which $\psi \rightarrow (Et \rightarrow \varphi)$ depends

Next we give the primary rule for relating the two types of knowledge.

R5. (Knowledge Relation)

From $S_i\varphi$ infer $C_i t$	where φ is an arbitrary sentence and t is any term occurring freely in φ

As an example, suppose Ed knows that $s = f(t_1, ..., t_n)$. Then, Ed can recognize all the arguments and the value of the function. It may seem to follow from this rule that all the subjects know all the words. By epistemic generalization, all subjects know all logical truths. And since $t = t$ is a logical truth, it would seem to follow from epistemic generalization and knowledge relation that all subjects know t. This appears to be disastrous since it means in particular that all subjects know all passwords and keys. One might argue that this is not a problem since no dangerous knowledge *about* keys and passwords derives from this. But we need not consider that because the derivation is flawed anyway. The rule of epistemic generalization says that if p is a theorem, then $S_i p$ is a theorem. But $t = t$ is not a theorem. $\forall x\ (x = x)$ is a theorem (in fact an axiom), but $t = t$ only follows from this provided that we also have Et. We must be careful to distinguish between $\forall x\ S_i(\varphi(x))$ and $S_i(\forall x\ \varphi(x))$.

Given the above discussion, it should be clear that we must reject the converse of the Barcan Formula (CBF), i.e., the conditional $S_i(\forall x\ \varphi(x)) \rightarrow \forall x\ S_i(\varphi(x))$. We must reject the Barcan Formula (BF) as well. Here is an example that illustrates why we must do so. Suppose that Ed knows all the words. Furthermore, suppose that for each word that he knows, he knows that he knows it. Then $\forall x\ S_{Ed}(C_{Ed}x)$ is true. It does not intuitively follow from this, however, that he knows that he knows all the words. For example, suppose that all words are passwords and that Ed has found all of them by searching in some way. This does not mean that he necessarily knows that he has now found them all and can stop searching. In other words $S_{Ed}(\forall x\ C_{Ed}x)$ does not intuitively follow from $\forall x\ S_{Ed}(C_{Ed}x)$. Thus both BF and CBF must be rejected in our system. Since unrestricted epistemic generalization leads to CBF we reject it in favor of the restricted version. And, the system also supports the reasoning in the above example; $\forall x\ S_{Ed}(C_{Ed}x)$ is true at w if Ed knows all the words at w, but for $S_{Ed}(\forall x\ C_{Ed}x)$ to be true at w, Ed would have to know all the words at all worlds accessible from w for him. Once we have shown soundness, the failure of BF follows.

Cryptographic Axioms

Before stating the cryptographic axioms, it will be useful to have a definition for the notion of an inverse key. This definition is completely eliminable and is made only for ease of notation and comprehension.

Definition of a (Two Sided) Key Inverse

$$\forall x, y\ [\ I(x, y) \leftrightarrow \forall z\ (d(y, e(x, z)) = z = e(x, d(y, z)))]$$

We are now in a position to state the two cryptographic axioms.

Secrecy Axiom

8.1. $\forall x, y, z, u\ [\ I(x, u) \wedge C_i u \wedge y = e(x, z) \wedge S_i(\exists x_1, x_2\ (y = e(x_1, x_2))) \rightarrow C_i z]$

Authenticity Axiom

8.2. $\forall x, y, z, u\ [\ I(x, u) \wedge C_i u \wedge y = d(x, z) \wedge S_i(\exists x_1, x_2\ (y = d(x_1, x_2))) \rightarrow C_i z]$

Obviously these axioms are intended to apply to an asymmetric (public key) cryptosystem. They apply equally well to a symmetric cryptosystem. In this case we simply have the add-

ed information that x = u, and we can drop the leftmost conjunct in the antecedent since it is always true.

The last conjunct in the antecedent of each axiom may seem unnecessary, especially since reasoners in this system are fairly idealized. Can we not just assume that, if i knows x and y, he will try to plug them into every formula at his disposal and see what results? Perhaps. But, even in this case it is clearer to be explicit about our idealized assumptions. Thus, if we wish to assume that i can always figure out that y is an encrypted word (when it is indeed an encrypted word), then we should do so explicitly. We should then assume the last conjunct as a premise rather than deleting it from the axiom.

METALOGIC

With our logic fully set out we can now begin our metalogical analysis. Actually we have already engaged in some analysis with our observations about the Barcan Formula and its converse. The first result we derive is the soundness of the logic. For the remainder of the report we adopt the following standard notational conventions. Let Γ stand for a finite set of sentences and φ, ψ, etc. stand for arbitrary sentences as before. '$\Gamma \vdash \varphi$' means that φ is derivable using the inference rules from Γ and the axioms. As usual, we follow the convention of writing '$\vdash \varphi$' for '$\Gamma \vdash \varphi$' when Γ consists solely of theorems. '$\Gamma \models \varphi$' means that, in all models, φ is true at all worlds where all the members of Γ are true.

Soundness

Theorem: If $\Gamma \vdash \varphi$, then $\Gamma \models \varphi$

To prove this we need the following lemma.

Lemma: All axioms are valid in all models provided that all freely occurring terms denote.

We assume that the lemma holds for axioms 1 through 5 since the proof is but a minor variation on the standard soundness result for **T**. (c.f. Hughes & Cresswell 1968 or Chellas 1980) Also, the result is trivial for the identity axioms, 6 and 7. So, all that remains is to prove the lemma for the cryptographic axioms, 8.1 and 8.2. Since the cases are very similar, we prove only that the secrecy axiom is valid in all models. First, note that the axiom cannot be undefined since it contains no free variables. If at some world w we instantiate x, y, z, u to t_1, t_2, t_3, and t_4 respectively, the resulting sentence is $I(t_1, t_4) \wedge C_i t_4 \wedge t_2 = e(t_1, t_3) \wedge S_i(\exists x_1, x_2\ (t_2 = e(x_1, x_2))) \rightarrow C_i t_3$, where $a(t_1), \ldots, a(t_4) \in d(w)$. Assume that the antecedent is true at w. Then, $t_4 \in [t_1^{-1}]$, and, at each world w′ accessible for i from w there exist some s_1, s_2 such that $a_{w'}(e(s_1, s_2)) = a(t_2) = a(e(t_1, t_3))$. These conditions are sufficient to guarantee that t_3 denotes at each such w′. Thus, $C_i t_3$ is true at w. So, the whole conditional is true at w, and, by universal generalization, 8.1 is true.

We now proceed to prove the theorem by showing that all the ways that φ can follow from Γ in a proof are ways that preserve validity.

Case i: φ is an axiom or member of Γ. Then $\Gamma \models \varphi$ trivially.

Case ii: φ is obtained by modus ponens from ψ and $\psi \rightarrow \varphi$. We proceed by strong induction. Suppose that soundness holds for all lines of a derivation up to φ. Then, by inductive hypothesis, $\Gamma \models \psi$ and $\Gamma \models \psi \rightarrow \varphi$. So, clearly $\Gamma \models \varphi$ by the definition of the assignment function.

Case iii: φ is obtained by epistemic generalization. Then φ is $S_i\psi$ for some ψ. Proceeding again by induction, we have $\Gamma \models \psi$. Since it must be the case that $\vdash \psi$, by inductive hypothesis $\models \psi$. So ψ is true in all worlds, hence true in all worlds accessible for i from any given world, i.e., $\models S_i\psi$. *Thus, a fortiori* $\Gamma \models S_i\psi$.

Case iv: φ is obtained by universal instantiation. Then, φ is of the form $\psi[t/x]$. Proceeding by induction, we assume $\Gamma \models \forall x\, \psi \wedge Et$. So, $\Gamma \models \forall x\, \psi$ and $\Gamma \models Et$. If x does not occur freely in ψ, *then* $\forall x\psi$ *is true iff* ψ *is true, and* ψ *is*$\psi[t/x]$ *in this case.* So $\Gamma \models \psi[t/x]$. If x does occur freely in ψ, then $\Gamma \models \psi[t/x]$ by the definition of the assignment function.

Case v: φ is obtained by universal generalization. So φ is of the form $\psi \rightarrow \forall x\, \theta[x/t]$, and, by inductive hypothesis, $\Gamma \models \psi \rightarrow Et \rightarrow \theta$ *where* t *is an arbitrary term not occurring freely in* ψ or any member of Γ. We may assume $\Gamma \models \psi$. (If ψ is false the result is trivial. And, if ψ is undefined, by inductive hypothesis all of Γ is undefined and again the result is trivial.) So, by definition of the assignment function, $\Gamma \models \psi \rightarrow \forall x\, \theta[x/t]$.

Case vi: φ is obtained by R5, knowledge relation. This rule can be seen to be valid simply by inspecting the assignment function.

QED

Completeness

Theorem: If $\Gamma \models \varphi$, then $\Gamma \vdash \varphi$.

We give a Henkin style proof for the completeness of the logic. That is, we construct a model where the worlds are maximal consistent sets of sentences and show that every consistent set is satisfiable. (It is a well known result that this is equivalent to completeness. For those unfamiliar with this, here is a very brief explanation. Restricting ourselves to the maximal consistent sets containing Γ, if $\Gamma \cup \{\varphi\}$ is valid in a set of worlds, then $\Gamma \cup \{\neg\, \varphi\}$ is not simultaneously satisfiable in any member of that set. Assuming Γ itself is consistent, if $\Gamma \cup \{\neg\, \varphi\}$ is inconsistent, it can only be because $\Gamma \vdash \varphi$. Thus, if we can prove that the inconsistency of $\Gamma \cup \{\neg\, \varphi\}$ follows from its failure to be simultaneously satisfiable, we will have shown completeness. We do this by proving the contrapositive—i.e., that every consistent set is satisfiable.)

We now take an *arbitrary consistent set of sentences and show that it is satisfiable.* Assume that we have a set of sentences Γ that is consistent with respect to the logic. By Lindenbaum's Lemma, this can be extended to a maximal consistent set v in some language L. Unfortunately, the basic Lindenbaum method does not guarantee quite enough. In order to prove what we want we must construct our maximal consistent sets so that the following condition is satisfied.

ω-completeness: If $w \vdash Et \rightarrow \varphi$ for every term t of L, then $w \vdash \forall x\, \varphi[x/t]$.

Note that this is equivalent to:

If $w \cup \{\neg \forall x \varphi\}$ is consistent, then for some term t of L, $w \cup \{\neg (Et \rightarrow \varphi[t/x])\}$ is consistent.

We say that an ω-complete, maximal consistent set of sentences of L is saturated (for L). To produce a saturated set, we proceed by Lindenbauming and show that our construction satisfies the equivalent formulation of ω-completeness. Begin with a consistent set Γ, with all sentences written in L. Order all sentences of L, $\{A_1, A_2, ...\}$. Then, define a series of sets $M_0 = \Gamma$, M_1, M_2, ... by letting $M_{i+1} = M_i \cup \{A_{i+1}\}$ if doing so leaves M_{i+1} consistent. Otherwise $M_{i+1} = M_i$. The union of the M_i's is maximally consistent by Lindenbaum's Lemma. To ensure ω-completeness we modify the construction slightly. If A_{i+1} is $\neg \forall x \varphi$ and $M_i \cup \{A_{i+1}\}$ is consistent, then we let $M_{i+1} = M_i \cup \{A_{i+1}, \neg (Et \rightarrow \varphi[t/x])\}$ where t is a term foreign to $M_i \cup \{A_{i+1}\}$. We claim that M_{i+1} is consistent if $M_i \cup \{A_{i+1}\}$ is consistent. If not, then it must be the case that $M_i \cup \{A_{i+1}\} \vdash Et \rightarrow \varphi[t/x]$. Since t does not occur in $M_i \cup \{A_{i+1}\}$, we can apply universal generalization to this in order to get that $M_i \cup \{A_{i+1}\} \vdash \forall x \varphi$. But then $M_i \cup \{A_{i+1}\}$ is inconsistent. Contradiction. This construction preserves consistency and guarantees both maximality and ω-completeness.

We now proceed to the construction of the standard model. Again, starting with a consistent set Γ of sentences of L, we extend this to a saturated set v by means of the above procedure. Now, consider a language L* containing infinitely more terms than L. We define the standard model $\langle W, R_1, ..., R_k, D, d, a\rangle$ as follows. Let W be the set of all sets w of sentences satisfying the following:

(1) Each world w is a saturated set for a language L_w, and L_w is a sublanguage of L* such that there are infinitely many terms of L* not occurring in L_w.

(2) $v \in W$.

(3) For all terms s and t which are members of both L_w and $L_{w'}$, $s = t \in w$ iff $s = t \in w'$.

(4) If $P(t_1, ...,t_n)$ is an expression of both L_w and $L_{w'}$, $P(t_1, ...,t_n) \in w$ iff $P(t_1, ...,t_n) \in w'$.

Clauses (3) and (4) require agreement between worlds with regard to the membership of certain sentences. Clause (2) is present simply to ensure that other worlds accommodate to v in such agreement.

The assignment function for terms is given by $a(t) = \{s: s = t \in \cup W\}$.

For an arbitrary n-ary predicate letter P the assignment function is given by
$a(P) = \{\langle t_1, ..., t_n\rangle: P(t_1, ..., t_n) \in \cup W\}$.

Definition of the assignment function for other arguments is as above.

The domain is given by $D = \bigcup_{w \in W} \{a(t): t \in L_w\}$, and thus $d(w) = \{a(t): t \in L_w\}$.

For each i, wR_iw' iff $S_i\varphi \in w \Rightarrow \varphi \in w'$.

With the specification of the standard model finished we proceed to the main step in our completeness proof, the truth lemma. Once the truth lemma is established we will have shown completeness since we will have shown that Γ (an arbitrary consistent set of sentences) is satisfied by the standard model.

Truth Lemma: If φ is a sentence of L_w, then $a_w(\varphi) = T$ iff $\varphi \in w$.

Case i: φ is of the form $\psi \wedge \theta$, $\psi \vee \theta$, $\psi \rightarrow \theta$, or $\neg\psi$. All of these follow by trivial inductive arguments.

Case ii: φ is of the form $s = t$. If $s = t \in w$, then $s \in L_w$ and $t \in L_w$. So, $a(s) \in d(w)$ and $a(t) \in d(w)$. We need that $a(s) = a(t)$. Suppose $u \in a(s)$. Then $u = s \in w'$ for some w'. Consider a language L_w+ formed by adding u to L_w (together with all resulting expressions). There exists a saturated set of sentences of L_w+, w^+ containing $u = t$. So, $a(s) = a(t)$ and $a_w(s = t) = T$. If $a_w(s = t) = T$, then $a(s), a(t) \in d(w)$ and $a(s) = a(t)$. So $s, t \in L_w$. Since w is maximal consistent, $s = t \in w$ or $s \neq t \in w$. But, if $a(s) = a(t)$, there is some world in W containing $s = t$. Thus, by clause (3) of the definition of W, $s = t \in w$.

Case iii: φ is of the form $C_i t$. If $C_i t \in w$, then, by maximal consistency, either $S_i(t = t) \in w$ or $\neg S_i(t = t) \in w$. We will see in *case v* below that if $\neg S_i(t = t) \in w$, then $t \neq t \in w'$ for some w' such that wR_iw', which is impossible. Thus, $S_i(t = t) \in w$. Therefore, $t = t \in w'$ for all w' such that wR_iw'. So, $a(t) \in d(w')$ for all w' such that wR_iw', and $a_w(C_i t) = T$. If $C_i t \notin w$, then, by knowledge relation and the maximal consistency of w, $S_i\psi \notin w$ for any sentence ψ containing any free occurrences of t. In particular, $S_i(t = t) \notin w$. And, as we have already mentioned, this leads to a contradiction.

Case iv: φ is of the form $P(t_1, ..., t_n)$. If $P(t_1, ..., t_n) \in w$, then $a(<t_1, ..., t_n>) \in a(P)$ and $a(t_1), ..., a(t_n) \in d(w)$. But, $a(<t_1, ..., t_n>) \in a(P)$ and $a(t_1), ..., a(t_n) \in d(w)$ iff $a_w(P(t_1, ..., t_n)) = T$. If $P(t_1, ..., t_n) \notin w$, then $\neg P(t_1, ..., t_n) \in w$. Thus, by clause (4) of the definition of W, $a(<t_1, ..., t_n>) \notin a(P)$. So, $a_w(P(t_1, ..., t_n)) \neq T$.

Case v: φ is of the form $S_i\psi$. If $S_i\psi \in w$, then $\psi \in w'$ for all w' such that wR_iw'. But, by inductive hypothesis, $\psi \in w'$ for all w' such that wR_iw' iff $a_{w'}(\psi) = T$ for all such w'. Thus, $a_{w'}(S_i\psi) = T$. If $S_i\psi \notin w$, then $\neg S_i\psi \in w$. We claim that if $\neg S_i\psi \in w$, then there is a $w' \in W$ such that wR_iw' and $\neg\psi \in w'$. To show this assume that $\neg S_i\psi \in w$ and let $\Delta = \{\varphi: S_i\varphi \in w\} \cup \{\neg\psi\}$. It is easy to see that Δ is consistent and contains only terms of L_w. It is not clear that there are infinitely many terms of L_w foreign to Δ. Thus, it is not clear that Δ can be extended to a saturated set of sentences for L_w. Let A be the set of terms occurring in L* but not in L_w. We can use A to extend Δ to a saturated set, but that set will not be in W because it will not have infinitely many terms of L* foreign to it. We solve this by partitioning A into two infinite sets A_1 and A_2. We then use A_1 to extend Δ to a saturated set w', and keep A_2 to ensure that there are infinitely many terms of L* foreign to $L_{w'}$. To establish the claim it remains only to show that wR_iw', but this follows trivially from the composition of Δ. With the claim thus shown, it follows by inductive hypothesis, that if $S_i\psi \notin w$, then $a_w(S_i\psi) \neq T$.

Case vi: φ is of the form $\forall x\ \psi$. By universal instantiation and ω-completeness, $\forall x\ \psi \in w$ is equivalent to $\psi[t/x] \in w$ for all t in L_w. But, by inductive hypothesis, this is equivalent to $a_w(\psi[t/x]) = T$ for all t in L_w. And, by the definition of d, this is equivalent to $a_w(\psi[t/x]) = T$ for all t such that $a(t) \in d(w)$, which is equivalent to $a_w(\forall x\ \psi) = T$.

QED

Corollary: The deduction theorem fails to hold for this logic.

This becomes obvious when we look at our rule of universal instantiation. From $\forall x\, \varphi \wedge Et$ we can infer $\varphi[t/x]$. However, from $\forall x\, \varphi$ we cannot infer $Et \rightarrow \varphi[t/x]$, for if t fails to denote, $Et \rightarrow \varphi[t/x]$ will have an undefined truth value. So, for example, $\forall x\, (x = x)$ is an axiom and thus true at all worlds. If the deduction theorem were to hold, then we could conclude from universal instantiation and completeness that $Et \rightarrow t = t$ is true at all worlds. But, this is clearly undefined at any world where t fails to denote. If we assume that all terms denote everywhere, we can prove a fairly standard first order deduction theorem. However, such a restriction would remove most—if not all—of the interesting innovations of our logic. Basically, the absence of a deduction theorem means that the logic does not have enough expressive power to capture its own consequence relation. While somewhat surprising there is no cause for concern, especially when we realize that this limitation applies only in those cases where one literally does not know what one is talking about. As mentioned above, ours is a logic of epistemic, not alethic, modalities.

CONCLUSIONS

In this report we have set out a logic and a formal semantics for that logic. We have subjected the logic to metalogical analysis. In particular, we have proven its soundness and completeness. While these are interesting results in their own right, they are especially important for logics that are to be applied to safety critical or security critical areas such as cryptographic protocols. Soundness and completeness do not guarantee that there will be no error in evaluating the security of a protocol. But, they do guarantee that there will be no formal error. Once we have formally specified a protocol, a logical derivation of any result concerning the specification will be correct—i.e. true of that specification—and anything that can be formally shown to be a semantic consequence of that specification will be provable in the logic. Of course, there is no guarantee that the specification is correct, but no logic can provide such a guarantee since this is not part of the formal analysis. And, it is only in the formal analysis that logic can hope to play a role.

Finally, we note that, although the logic has been devised specifically as a logic for cryptographic protocol analysis, its ability to represent knowledge in the sense of familiarity is clearly applicable in other contexts. How this and other unique features of the logic might be applied, and in which contexts, is an interesting area for further study.

ACKNOWLEDGEMENTS

I am grateful to Jim Gray, John McLean, and Cathy Meadows for their numerous helpful suggestions and comments on a draft of this report.

REFERENCES

J. Barwise, "On the Model Theory of Common Knowledge," in *The Situation in Logic* (CSLI, Stanford, 1989), Ch. 9, pp. 201—220.

E. Bencivenga, "Free Logics," in *Handbook of Philosophical Logic, vol. III: Alternatives to Classical Logic* (D. Reidel Publishing Co., Dordrecht, 1986), pp. 373—426.

B.F. Chellas, *Modal Logic: An Introduction* (Cambridge University Press, Cambridge, 1980).

R. Eberle, "A Logic of Believing, Knowing, and Inferring," *Synthese*, **26**, 356—382 (1974).

R. Fagin and J. Halpern, "Belief, Awareness, and Limited Reasoning: Preliminary Report," in *Proc. of the Ninth International Joint Conference on Artificial Intelligence, vol. 1* (William Kaufmann Inc., Los Altos, Calif., 1985), pp. 491—501.

J. Garson, "Quantification in Modal Logic," in *Handbook of Philosophical Logic, vol. II: Extensions of Classical Logic*, (D. Reidel Publishing Co., Dordrecht, 1984), pp. 249—308.

S. Goldwasser, S. Micali, and C. Rackoff, "The Knowledge Complexity of Interactive Proof Systems," in *Proceedings of the Seventeenth ACM Symposium on the Theory of Computing* (ACM Press, New York, 1985), pp. 291—304.

J. Halpern and M. Vardi, 'The Complexity of Reasoning About Knowledge and Time: Extended Abstract," in *Proceedings of the Eighteenth Annual Symposium on the Theory of Computing* (ACM Press, New York, 1986), pp. 304—315.

J. Hintikka, *Knowledge and Belief: An Introduction to the Logic of Two Notions* (Cornell University Press, Ithaca, 1962).

G.E. Hughes and M.J. Cresswell, *An Introduction to Modal Logic* (Methuen and Co., London, 1968).

R. Kemmerer, "Using Formal Verification Techniques to Analyze Encryption Protocols," in *Security and Privacy, vol. 2: Proceedings of the 6^{th}, 7^{th}, & 8^{th} Symposia, 1985—1987* (IEEE Computer Society Press, Los Alamitos, Calif., 1987), pp. 134—139.

H. Levesque, "A Logic of Implicit and Explicit Belief," in *AAAI—84, Proceedings of the National Conference on Artificial Intelligence* (William Kaufmann Inc., Los Altos, Calif., 1983), pp. 198—202.

L. Moser, "A Logic of Knowledge and Belief for Reasoning about Computer Security," in *Proceedings of the Computer Security Foundations Workshop II* (IEEE Computer Society Press, Los Alamitos, Calif., 1989), pp. 57—63.

P. Syverson, "Formal Semantics for Logics of Cryptographic Protocols," in *Proceedings of the Computer Security Foundations Workshop III* (IEEE Computer Society Press, Los Alamitos, Calif., 1989), pp. 32—41.

P. Syverson, "The Use of Logic in the Analysis of Cryptographic Protocols," submitted for publication.